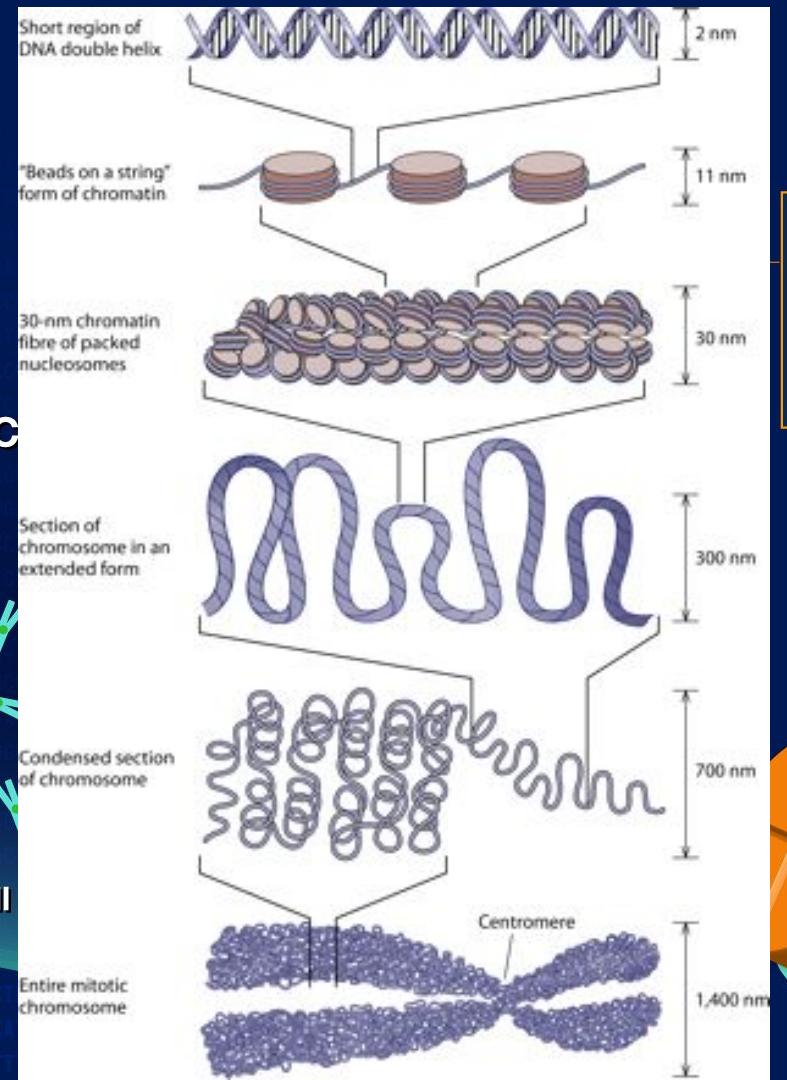


What are good drug targets and how to find them?

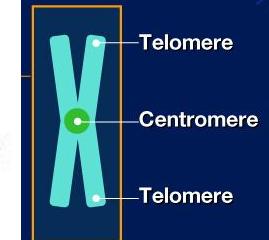
*Mathematical and Computational Biology in Drug
Discovery (MCBDD), Module I*

Dr. Jitao David Zhang, March 2022

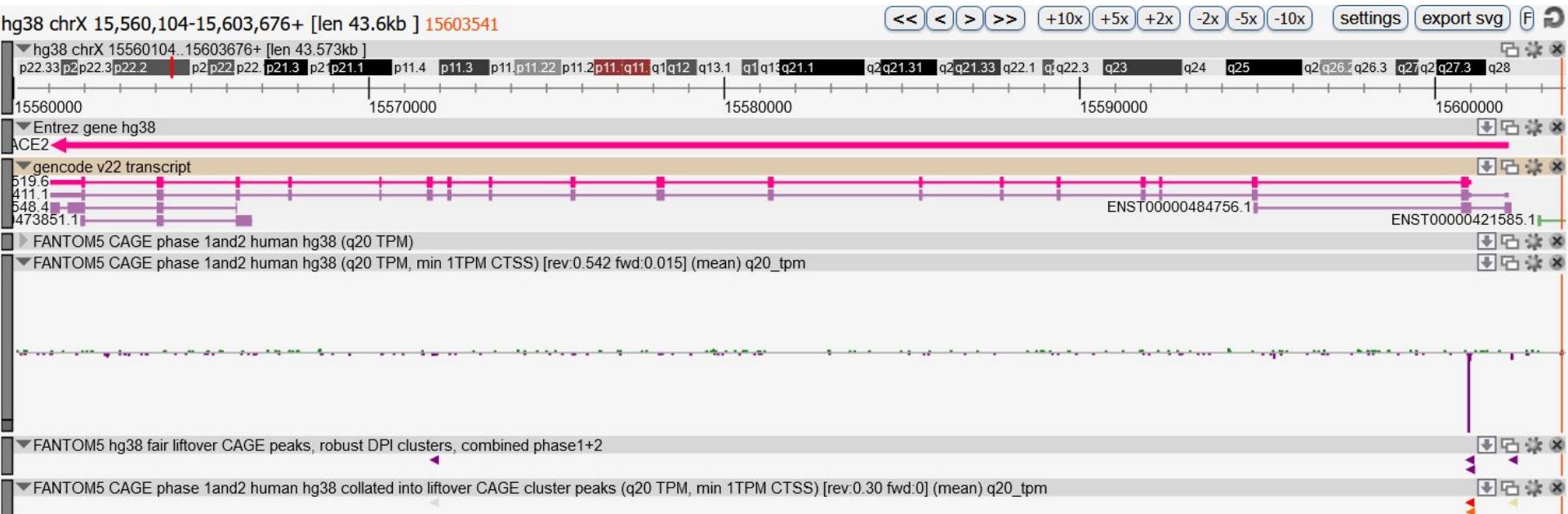
Chromosome



NHGRI FACT SHEETS
genome.gov



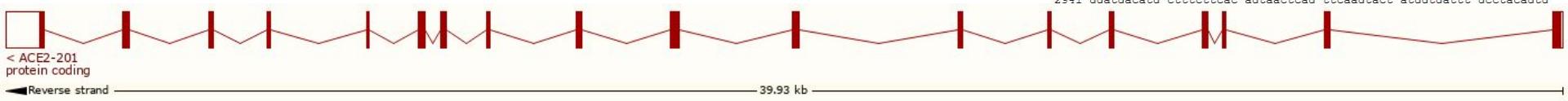
Gene structure and gene expression



ACE2 viewed in **FANTOM5/ZENBU**

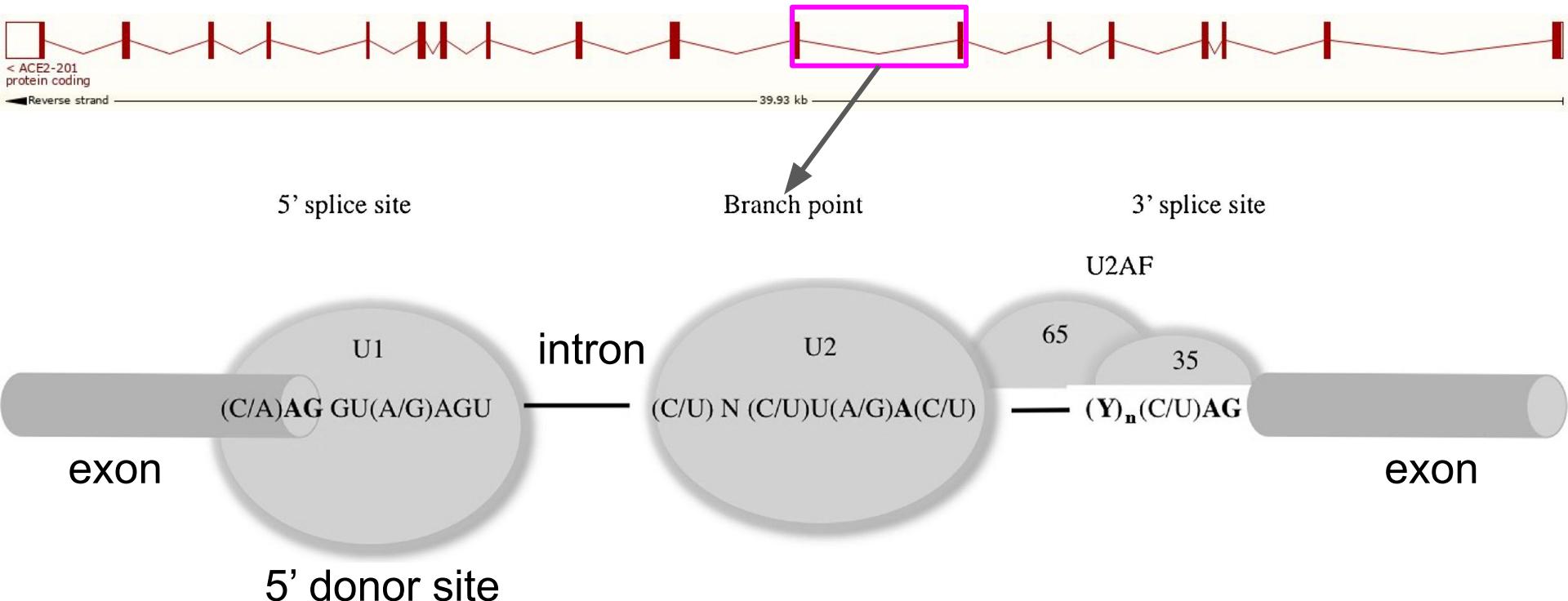
A mRNA of ACE2

- RefSeq record [NM_001371415.1](#)
- EnsEMBL record
[ENST00000252519.8](#)
- Red and blue boxes: start codon (ATG) and stop codon (TAG). The region between them is called the *coding sequence* (CDS).



1 agtcttaggaa aagtcatcca gtggatgtga tcttggctca caggggcac **a tg** caagctc
61 61 ttccctggc tccttcagcc ttgttgctgt aactgtgtc cagttccactt cttggggaca
121 121 ggccaaagaca tttttggaca agtttacca acggaaacaa gacctgttct atccaaatgtc
181 181 acttgtctt tggtttttata acaccaatata tactgaagag aatgttccaa acatgttata
241 241 tgctggggcac aaatgtgtc ctttttttata ggaaacatgc acatctggcc aatgttatcc
301 301 ggtttttttata atccaaatata tcacaaatgc gttttttata ctttgttctt ctttgttctt
361 361 gtcttcaggta ctccatggaa acaaagacaa acgggttccaa acatgttata atccaaatgtc
421 421 caccatcttc atgatgtggaa aagtttgttata cccagataat ccacaaagaat gttttttata
481 481 tgaaccatgttata tggtttttata taatggccaa cttgttacatgc tttttttata
541 541 ttggggaaatgg tggatgttccaa atccaaatata gttttttata ctttgttctt
601 601 ggtttttata aatgttccaa acgggttccaa acatgttata gttttttata
661 661 aggagactat gaatgtttata gggatgttccaa ctatgttccaa acgggttccaa
721 721 agatgtggaa cttttttata aatccaaatata accattttata gttttttata
781 781 gggggccaaatgg ttgtatgttccaa cttttttata ctatgttccaa
841 841 tttttttata gttttttata gttttttata gttttttata
901 901 tttttttata gttttttata gttttttata gttttttata
961 961 acaccaatata tttttttata gttttttata gttttttata
1021 1021 tttttttata gttttttata gttttttata gttttttata
1081 1081 cccatccatgc gttttttata gttttttata gttttttata
1141 1141 gaaatgttccaa gttttttata gttttttata gttttttata
1201 1201 atatgttccaa gttttttata gttttttata gttttttata
1261 1261 tggggaaatgg tttttttata gttttttata gttttttata
1321 1321 gttttttata gttttttata gttttttata gttttttata
1381 1381 cacgtttttata gttttttata gttttttata gttttttata
1441 1441 taaatggggaa atatccaaatata accatgttccaa
1501 1501 agttttttata gttttttata gttttttata gttttttata
1561 1561 coatgtttttata aatgttccaa cttttttata gttttttata
1621 1621 gttttttata gttttttata gttttttata gttttttata
1681 1681 ctatgtttttata gttttttata gttttttata gttttttata
1741 1741 accctggaccatgc tttttttata gttttttata gttttttata
1801 1801 gttttttata gttttttata gttttttata gttttttata
1861 1861 gggatgttccaa atccaaatata gttttttata gttttttata
1921 1921 aatccaaatata gttttttata gttttttata gttttttata
1981 1981 atatgttccaa gttttttata gttttttata gttttttata
2041 2041 tttttttata gttttttata gttttttata gttttttata
2101 2101 tttttttata gttttttata gttttttata gttttttata
2161 2161 ctttgttccaa atccaaatata gttttttata gttttttata
2221 2221 tttttttata gttttttata gttttttata gttttttata
2281 2281 tttttttata gttttttata gttttttata gttttttata
2341 2341 gttttttata gttttttata gttttttata gttttttata
2401 2401 ctttgttccaa atccaaatata gttttttata gttttttata
2461 2461 ctttgttccaa atccaaatata gttttttata gttttttata
2521 2521 tttttttata gttttttata gttttttata gttttttata
2581 2581 tttttttata gttttttata gttttttata gttttttata
2641 2641 gttttttata gttttttata gttttttata gttttttata
2701 2701 tttttttata gttttttata gttttttata gttttttata
2761 2761 tttttttata gttttttata gttttttata gttttttata
2821 2821 tttttttata gttttttata gttttttata gttttttata
2881 2881 tttttttata gttttttata gttttttata gttttttata
2941 2941 tttttttata gttttttata gttttttata gttttttata

The splicing code



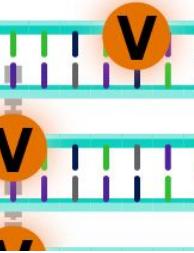
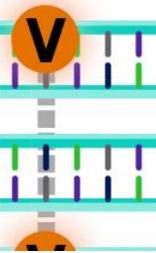
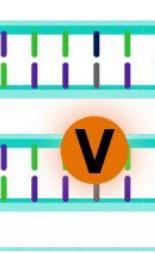
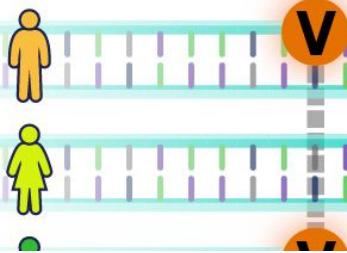


Person one

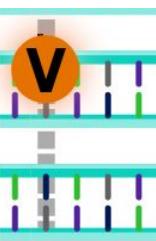
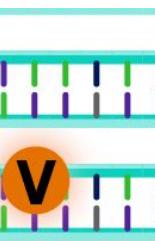
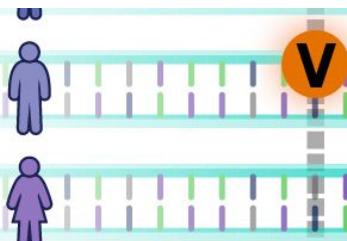
A G A C G C T

Variant ID	Source	HGVS Consequence	VEP Annotation	LoF Curation	Clinical Significance	Flags	Allele Count
17-7579017-C-T	E	c.74+22G>A	● intron		Likely benign		1
17-7579831-C-T	E	c.74+8G>A	● splice region		Likely benign		1
17-7579924-G-A	E G	c.-12C>T	● 5' UTR		Likely benign		7
17-7579932-G-C	E	c.-20C>G	● 5' UTR		Likely benign		2
17-7578142-C-A	E G	c.672+35G>T	● intron		not provided		9
17-7577142-C-A	E	p.Gly266Ter	● stop gained		Pathogenic		1
17-7578188-C-A	E	p.Glu221Ter	● stop gained		Pathogenic		1
17-7578263-G-A	E	p.Arg196Ter	● stop gained		Pathogenic		1
17-7576928-TAGGAA...	E	c.920-14_920-3delTGC...	● splice region		Uncertain significance		2
17-7578171-C-A	E G	c.672+6G>T	● splice region		Uncertain significance		2
17-7578171-C-T	E	c.672+6G>A	● splice region		Uncertain significance		1
17-7579934-C-T	G	c.-22G>A	● 5' UTR		Uncertain significance		1
17-7565206-T-A	G	c.*51A>T †	● 3' UTR				1
17-7565222-C-T	G	c.*35G>A †	● 3' UTR				1

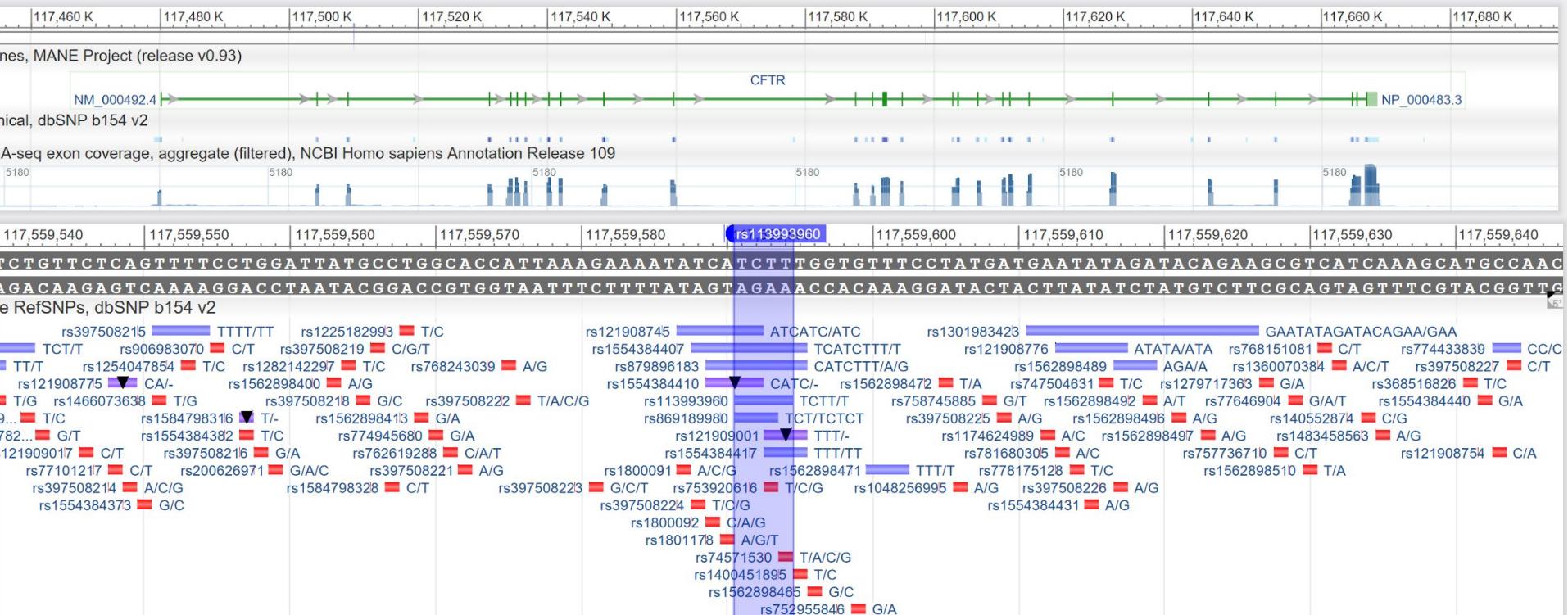



a
CNV
Other SV (non-CNV)
Unresolved

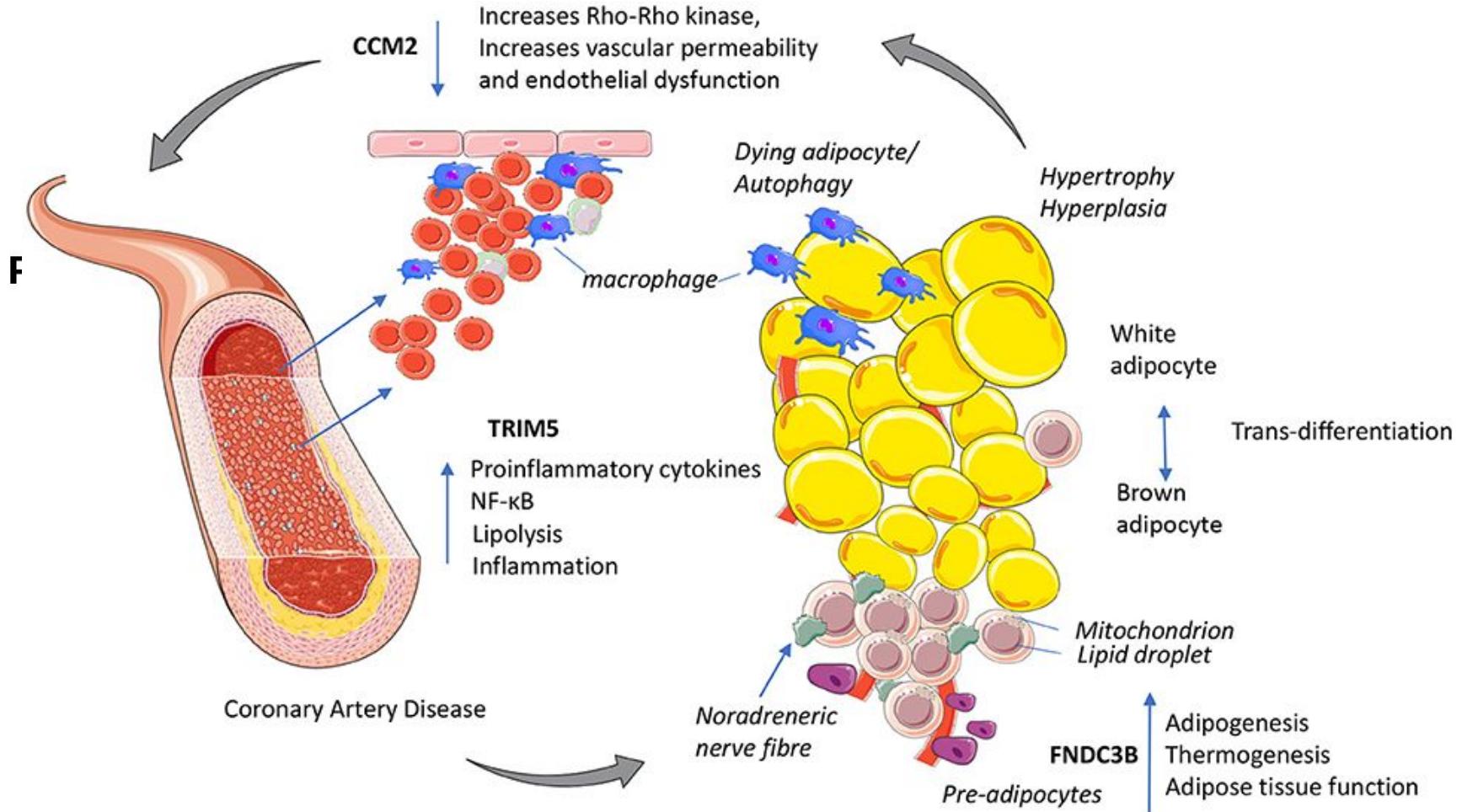
SV class	Deletion	Duplication	Multiallelic CNV	Insertion	Inversion	Translocation	Complex SV	Breakends
Abbrev.	-DEL	-DUP	-MCNV	-INS	-INV	-CTX	-CPX	-BND
Ref.								
Example alternatives								
							(See Fig. 2)	Discarded



Cystic fibrosis

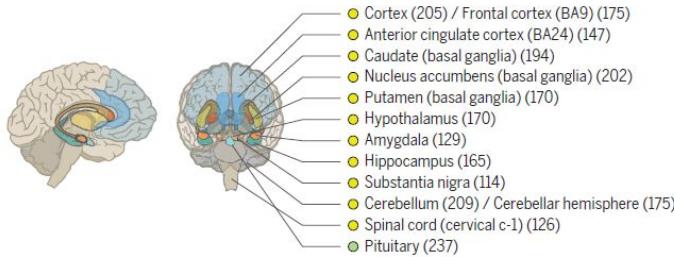


The *CFTR* gene (Chr 7), and rs113993960, the most common cause of CF



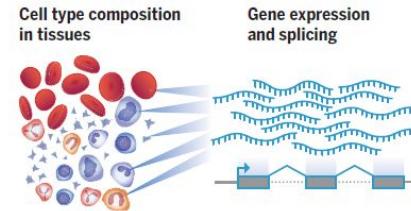
Dna Se

A



B

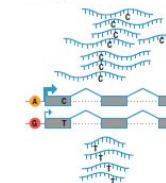
Cell type composition
in tissues



Gene expression
and splicing

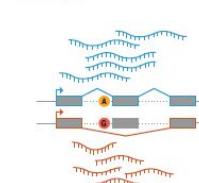
Expression quantitative
trait loci (eQTLs)

cis-eQTLs

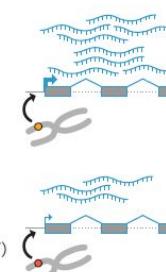


Splicing quantitative
trait loci (sQTLs)

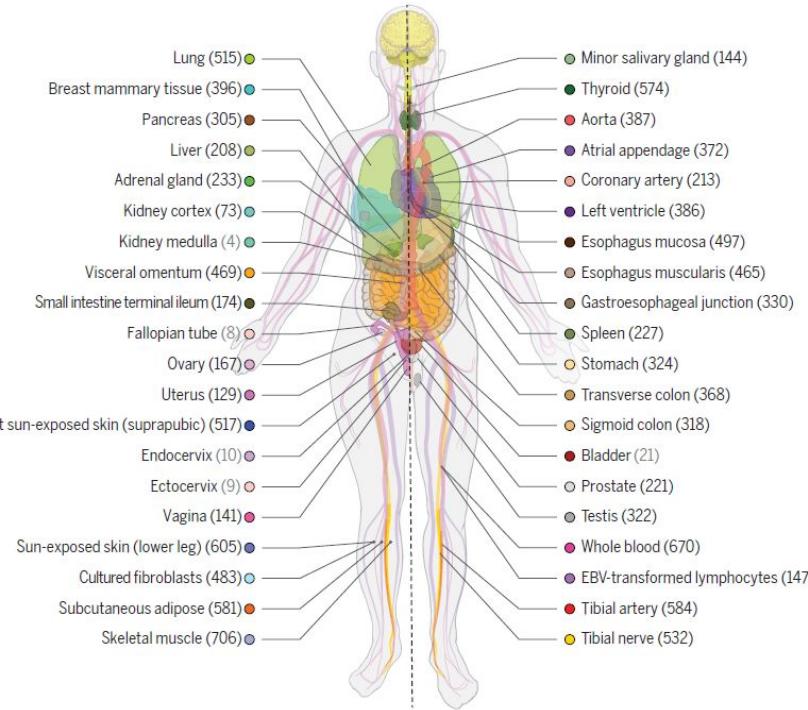
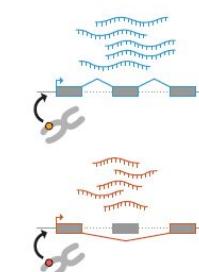
cis-sQTLs



trans-eQTLs



trans-sQTLs



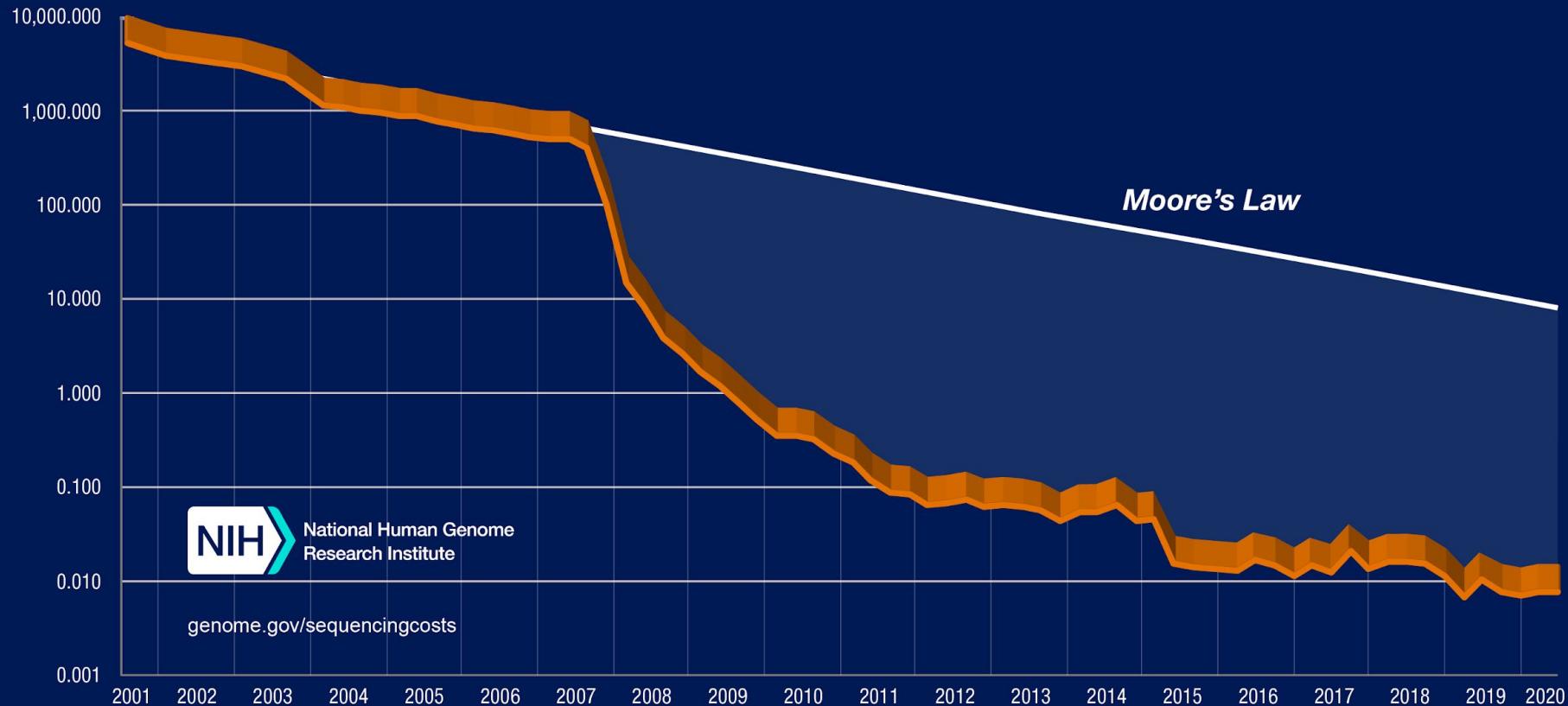
FACT SHEETS
genome.gov

GTEX (v8)



National Human
Genome Research Institute

Cost per Raw Megabase of DNA Sequence



National Human Genome
Research Institute

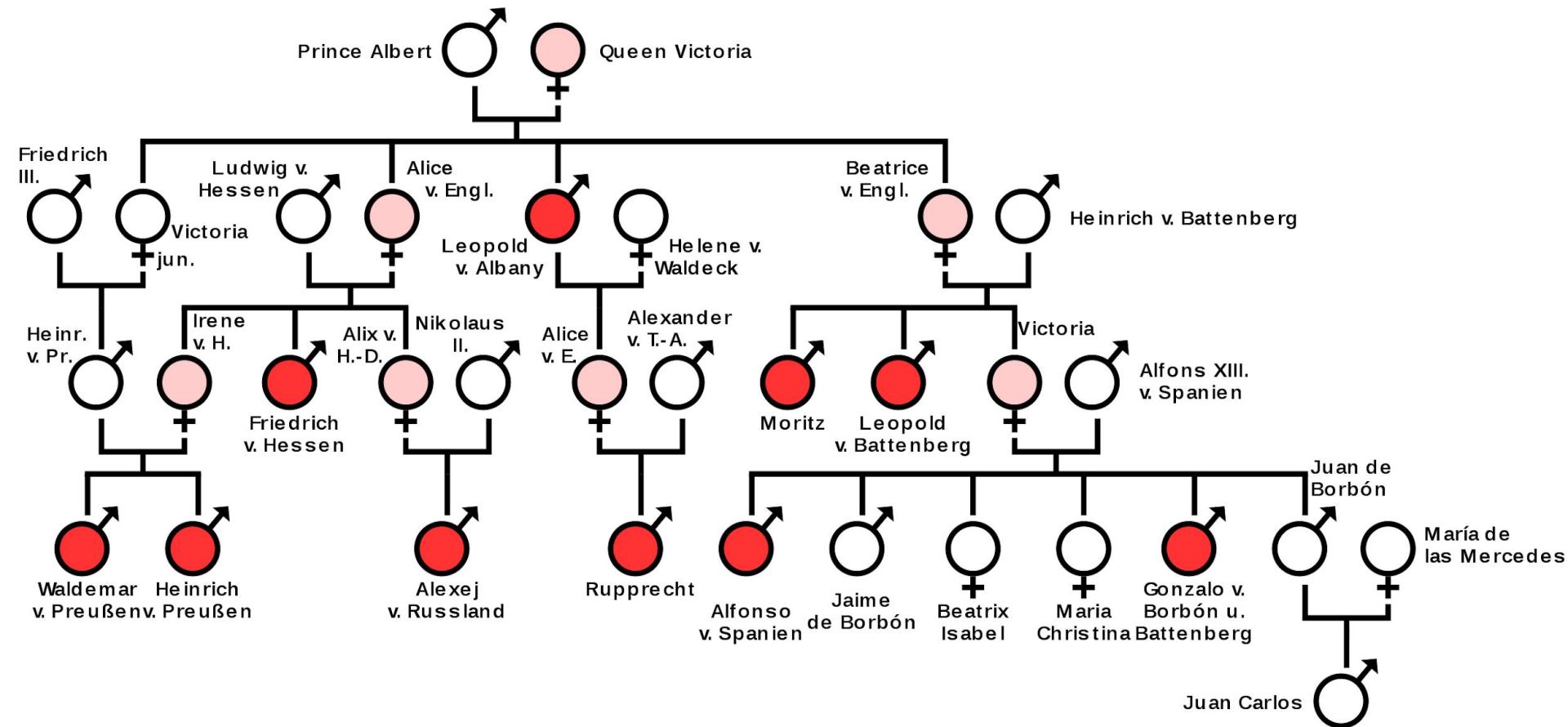
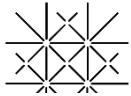
genome.gov/sequencingcosts

Cost per Human Genome



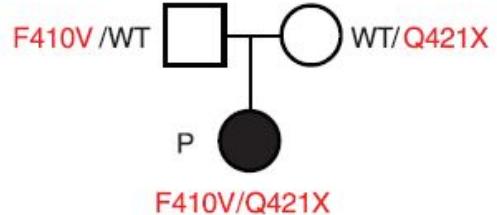


Haemophilia in the descendants of Queen Victoria

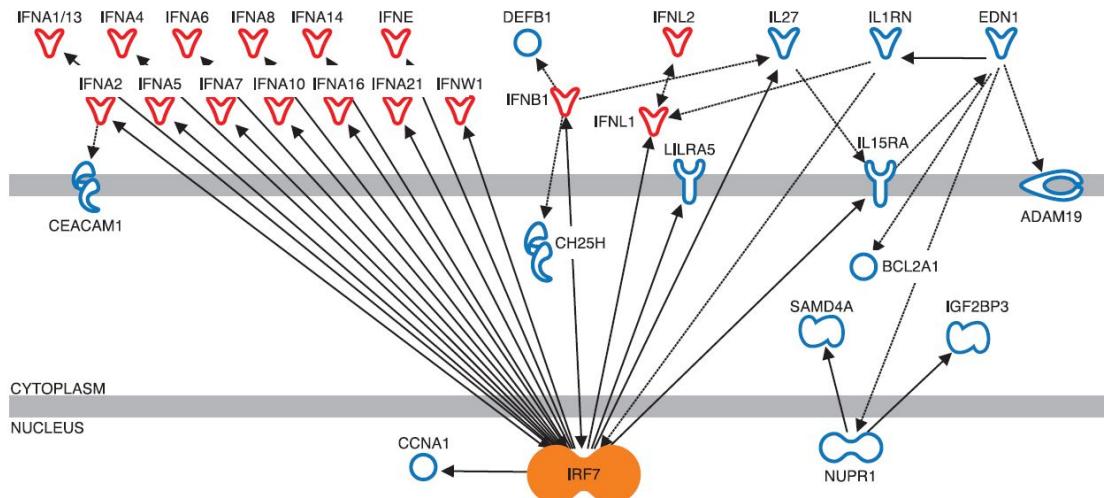
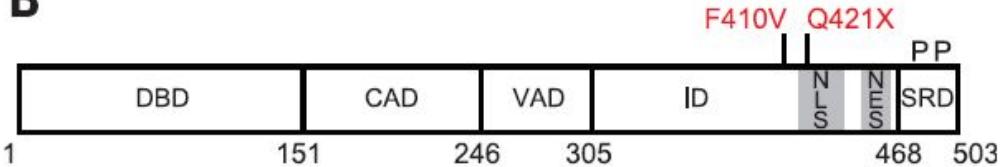


Life-threatening influenza infection in human IRF7 deficiency detected by trio sequencing

A



B



Exercise of *inference* (I)

The company *Fränzi and Friends* developed a new quick test at home for SARS-Cov-2 which is pending regulatory agency's review. The test has been shown to have a sensitivity of 99% and a specificity of 99%.

Suppose that Fred uses the test by *Fränzi and Friends* and the test was positive. Assume that 5% of the population is in fact infected. Was is your guess about the probability that Fred is indeed infected?

(Sensitivity is predicted true positive divided by all true positive; specificity is predicted true negative divided by all true negative).

Exercise of *inference* (I) - variants

The company *Fränzi and Friends* developed a new quick test at home for SARS-Cov-2 which is pending regulatory agency's review. When test with 100 SARS-Cov-2 patients, 99 report positive and one reports negative. When test with 100 healthy volunteers, 99 report negative and one reports positive.

Suppose that Fred uses the test by *Fränzi and Friends* and the test was positive. There are 30,000 people in the city where Fred lives; among them 1,500 are infected with SARS-Cov-2. What is the likelihood that Fred is truly infected given his positive test?

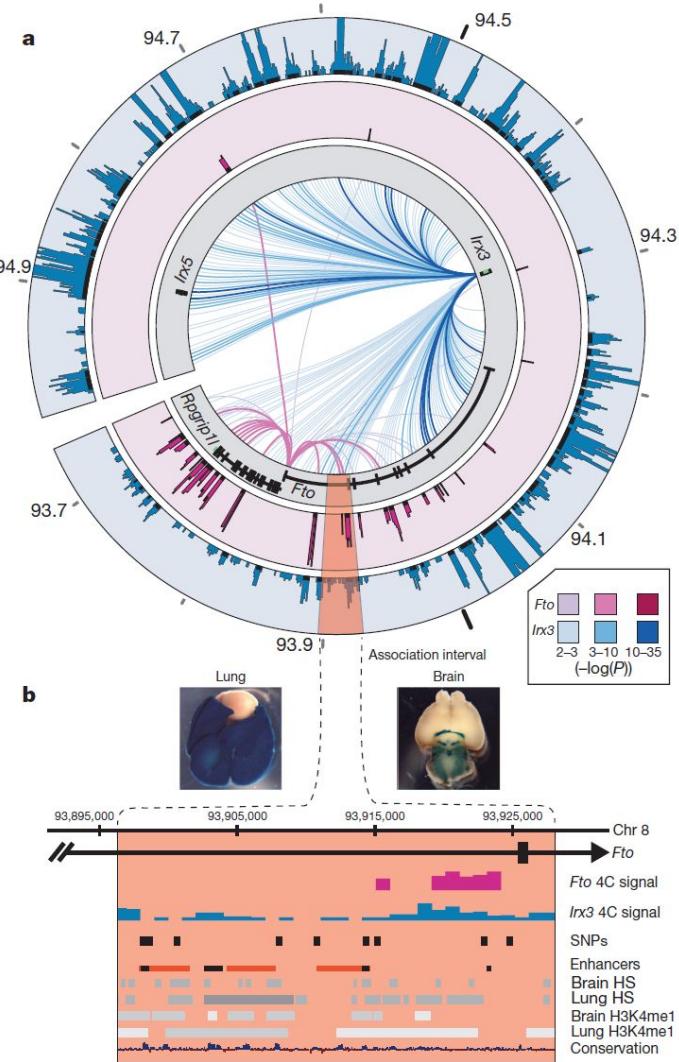
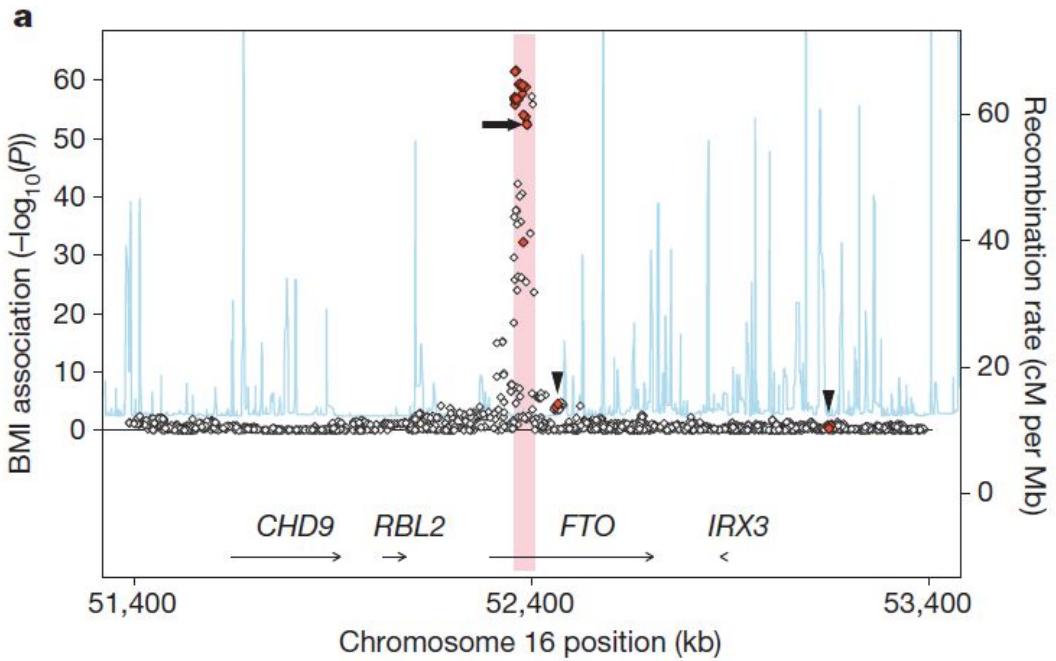
Exercise of *inference* (II)

Let's play a game! (The credit goes to David MacKay)

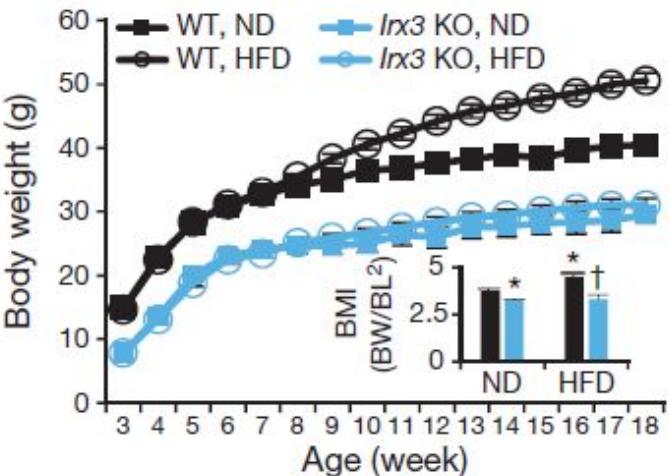
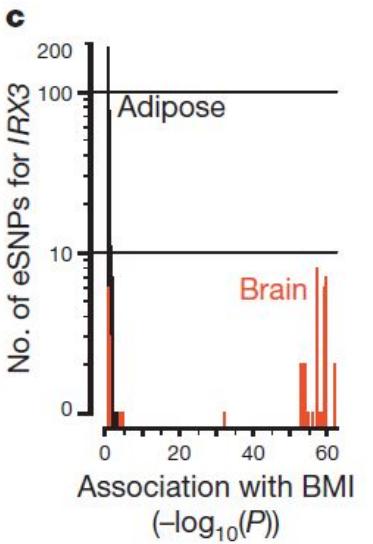
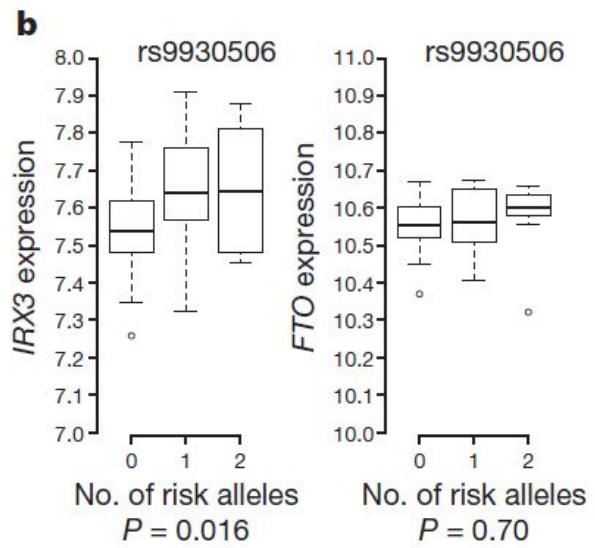
Exercise of *inference* (III)

- Cao and Moult (BMC Genomics, 2014) reported studied overlap between drug targets and GWAS hits.
- Use the data in the Table 1 of the paper ([cloned here](#)) to answer following the following two questions:
 - a. Assuming we know *nothing* about a gene (let's call it gene *WKN1*), what is the probability that the gene is a target for a disease listed here?
 - b. Assuming that we know nothing about another gene *WKN2* but that it is a GWAS hit for a disease, what is the probability that *WKN2* is a target for that disease?

Is FTO a good target for obesity?



If at all, **IRX3** is a more probable target



Recap of the biology we talked about last time



ACE2 viewed in NCBI Genome Browser

The Human
Genome
and
Variations

Gene
Structure
and gene
expression

DNA and
RNA
sequencing

Recap of the math we talked about last time

- ***Always write down your probabilities.***
- The probability theory and the Bayes theorem for inference.

	Sensitivity	Specificity
NPS	86% (90% CI 77–93)	99.93% (90% CI 99.77–99.99)
Saliva	92% (90% CI 83–97)	99.96% (90%CI 99.85–100)

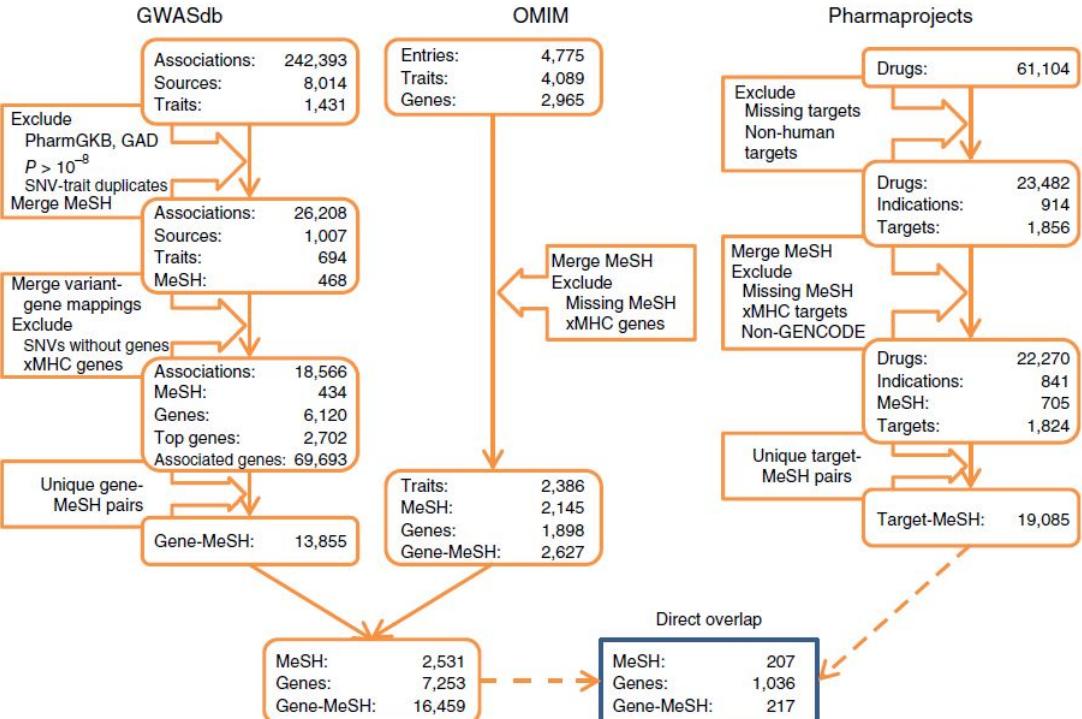
NPS = nasopharyngeal swabs. [Yokota et al., 2020](#)

- A list of approved serological tests and their accuracy can be found [here](#);
- *Does He Have It?* by Bill Casselman (AMS)

Other questions

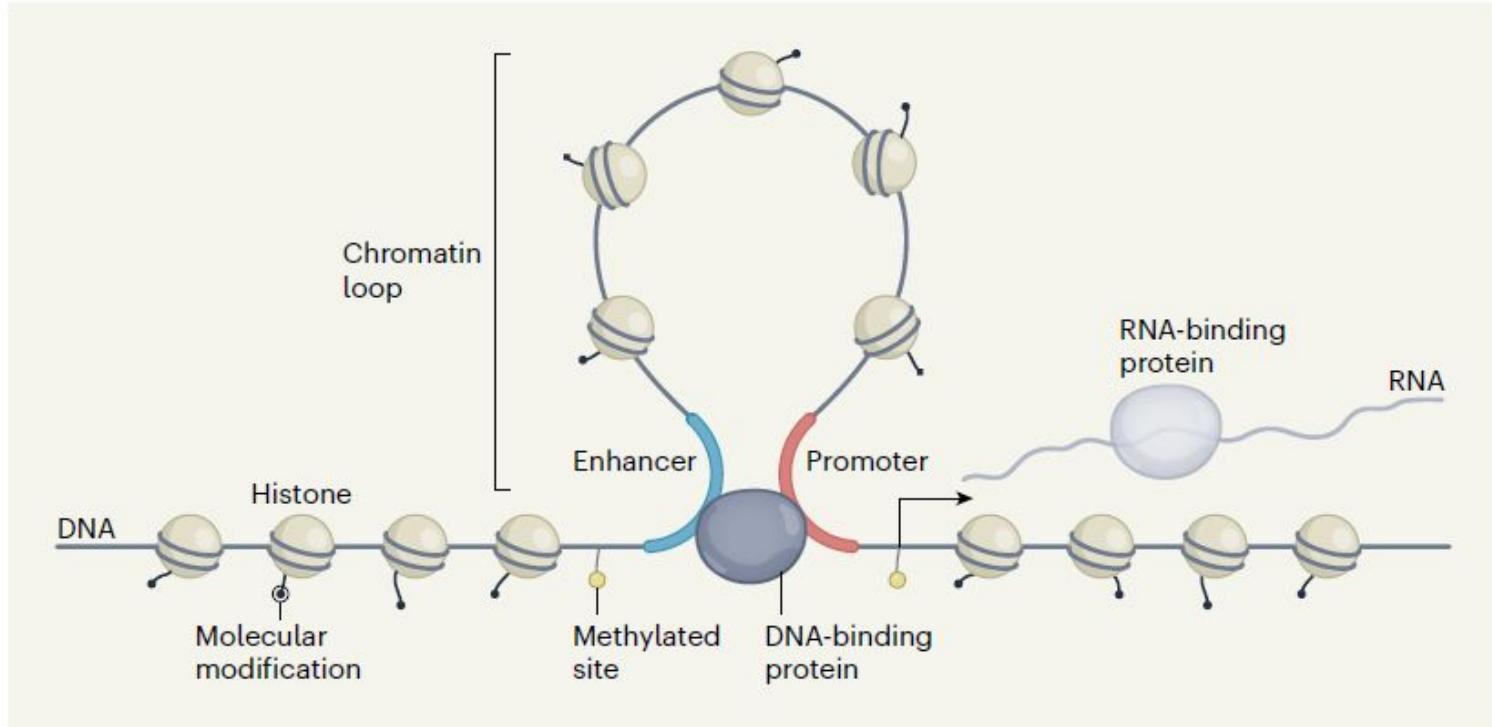
- Why I recommended the GOT-IT paper? How can academia and industry work together towards good targets?
- Did sequencing cost always follow the Moore's law?
- What happens if there are ATGs (AUGs in RNA) in 5'-untranslated region?
 - In some cases, there are alternative start codons;
 - However, in most cases, the ATGs in 5'-untranslated region seem to be always ignored by the translational machinery. A study (Rogozin, Bioinformatics, 2001) suggested that those AUGs may ensure low basal expression and generate regulatory elements.
- Which target level (gene or protein) is more useful for the target identification?
In the lecture, gene-level approach seems not so promising.
- Is blue eyeness a marker of Neanderthalian origin? *The story of OAC2*

Impact by a factor of ~2 estimated by Nelson et al.

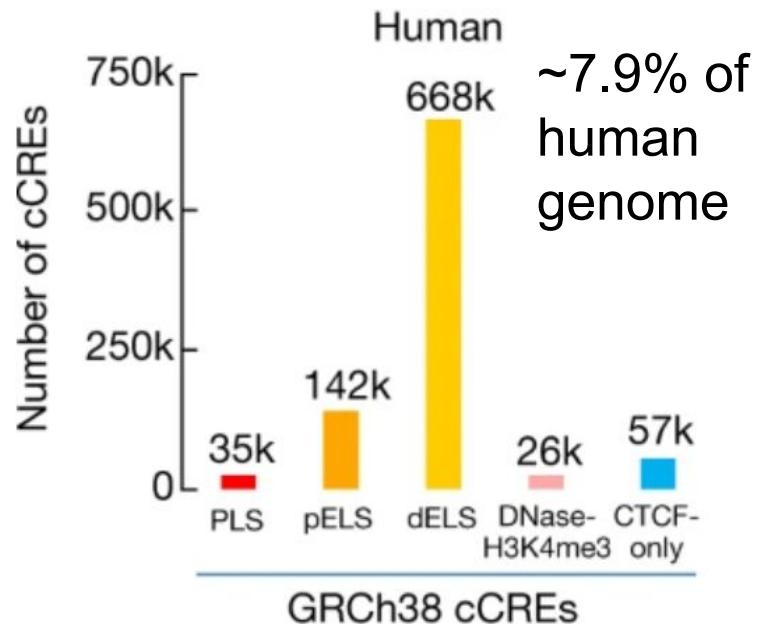
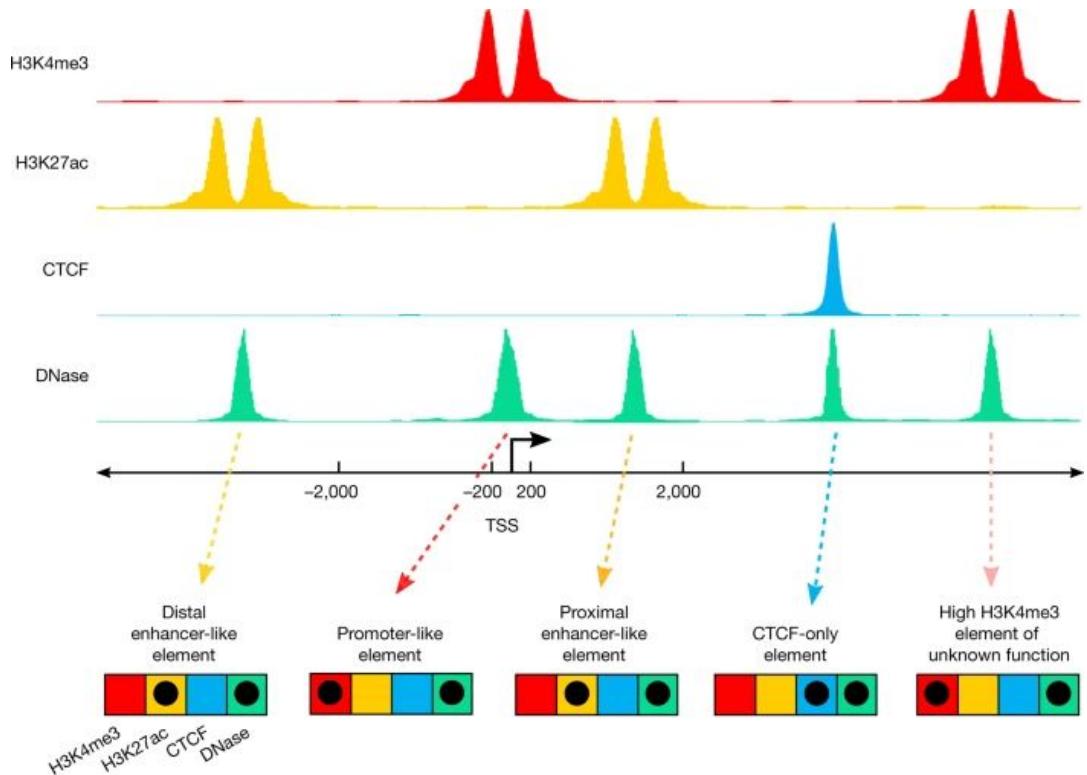


Disease \longleftrightarrow Gene \longleftrightarrow Drug

Much of the non-coding genome is junk, some is regulatory



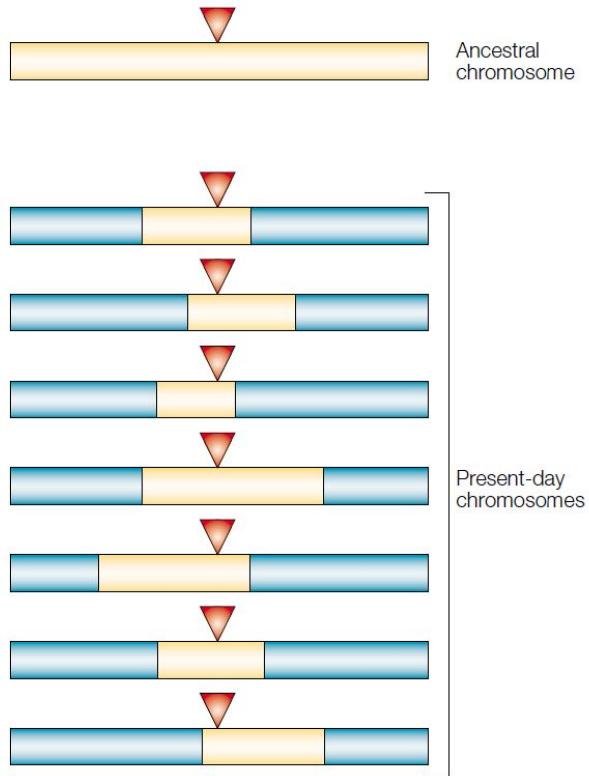
Candidate cis-regulatory elements (cCRE)



<https://screen.encodeproject.org/>

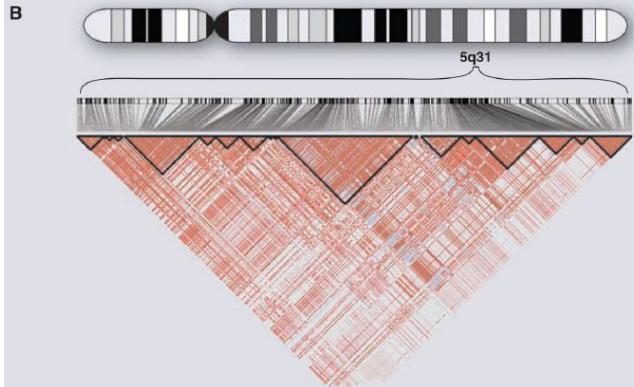
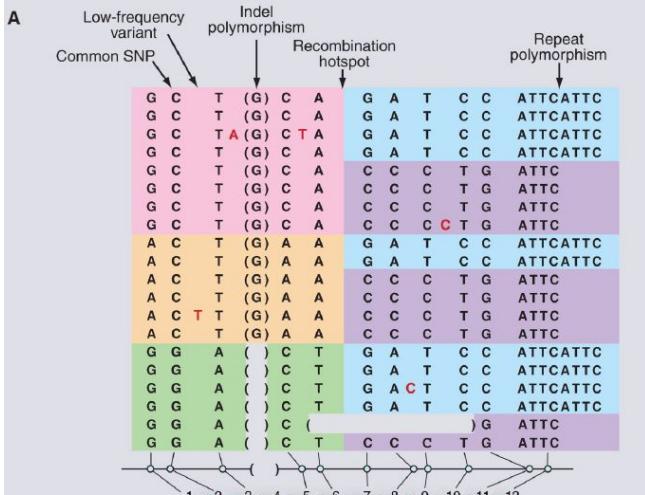
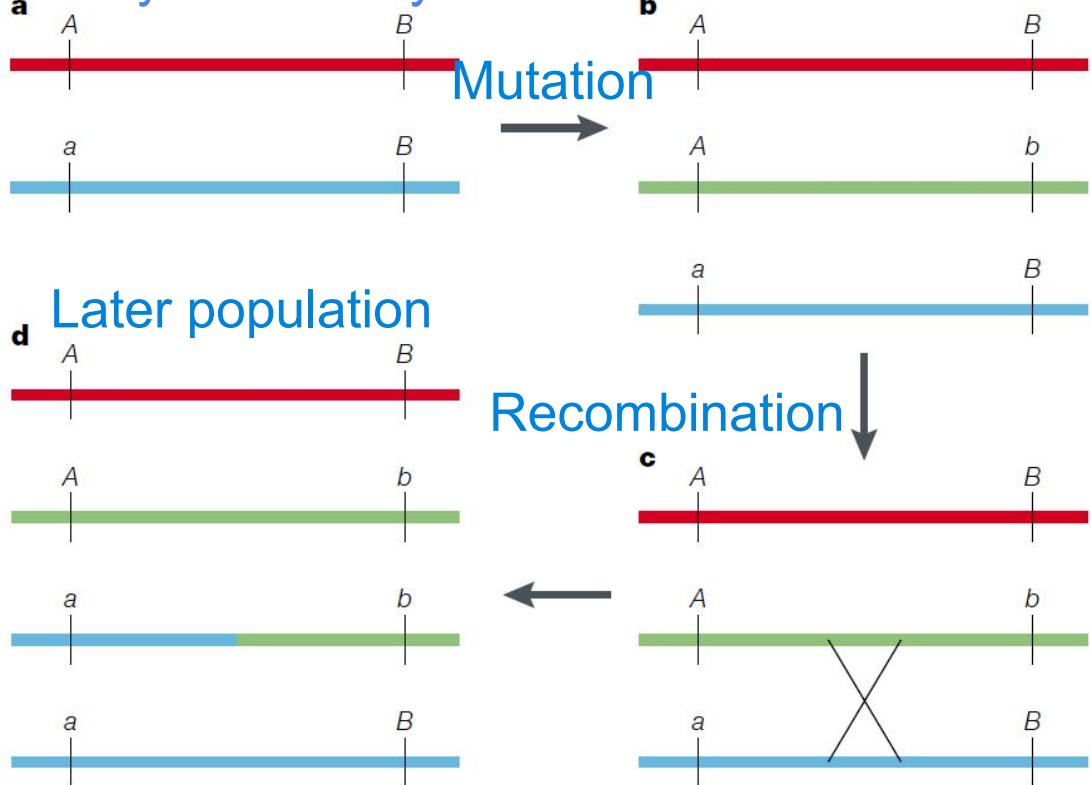
Linkage Disequilibrium in human genome

Particular alleles (single gene copies) at neighbouring loci tend to be co-inherited. For tightly linked loci, this might lead to associations between alleles in the population. This property is known as linkage disequilibrium (LD).

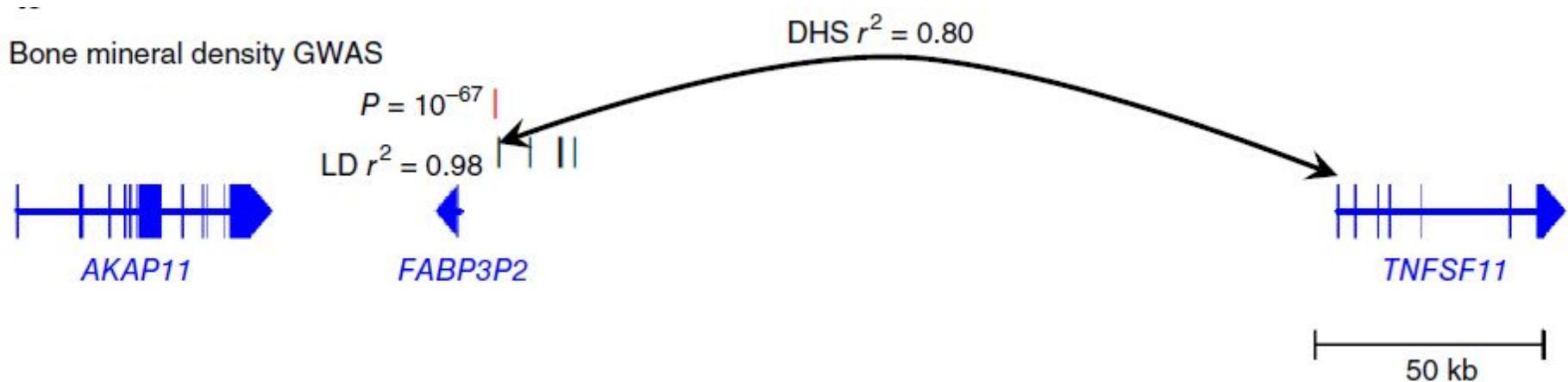


Population genetics helps with disease mapping

Early in ancestry



Correlation between DNase I hypersensitive (DHS) sites helps linking genetic variants with genes



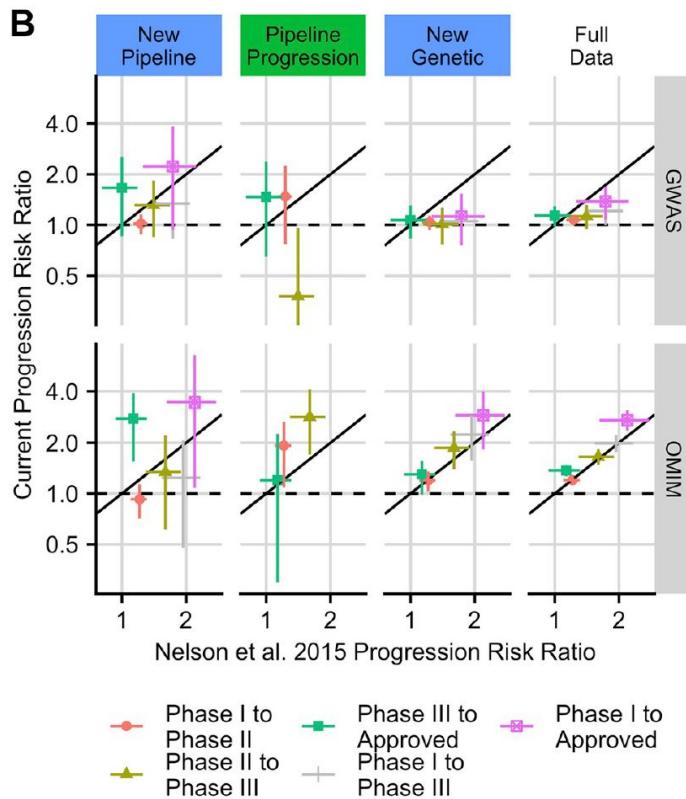
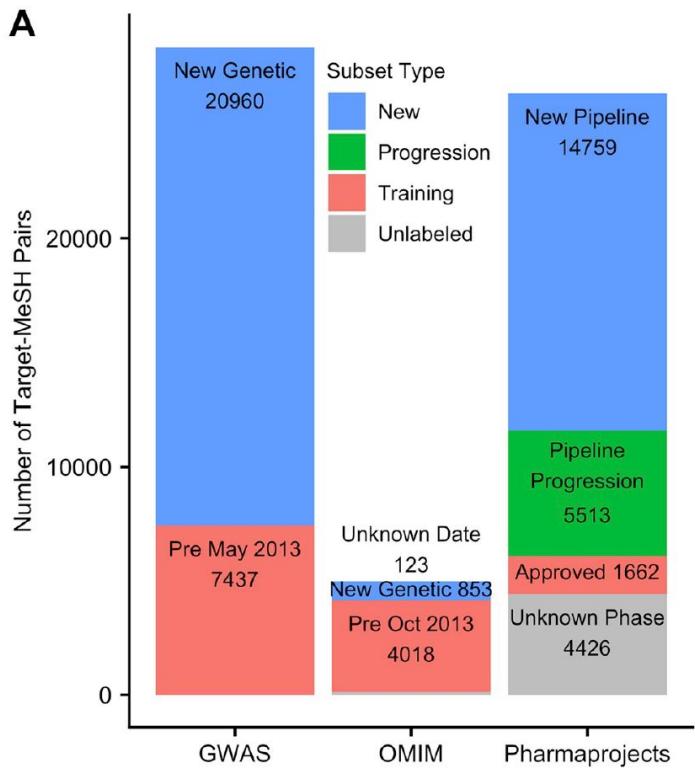
Nelson *et al.* inferred that genetic support offers an likelihood ratio ~2

Table 1 The relative value of genetic support for the probability that a target-indication pair progresses along the drug development pipeline, based on historical drug trial information

Progression	$p(\text{progress} \text{genetic support}) / (\text{progress} \text{no genetic support})$		
	GWASdb and OMIM	GWASdb	OMIM
Phase I to phase II	1.2 (1.1–1.3)	1.2 (1.1–1.3)	1.2 (1.1–1.3)
Phase II to phase III	1.5 (1.3–1.7)	1.4 (1.2–1.7)	1.6 (1.3–1.9)
Phase III to approval	1.1 (1.0–1.2)	1.0 (0.8–1.2)	1.1 (0.9–1.3)
Phase I to phase III	1.8 (1.5–2.1)	1.8 (1.4–2.1)	1.9 (1.5–2.3)
Phase I to approval	2.0 (1.6–2.4)	1.8 (1.3–2.3)	2.2 (1.6–2.8)

Values give the ratio of the probability of a target-indication pair progressing given genetic support to the probability of progressing without genetic support; 95% confidence intervals are given in parentheses.

Follow-up study by King et al., 2019



Genes with *biologically understandable* genetic association are more likely to be good targets

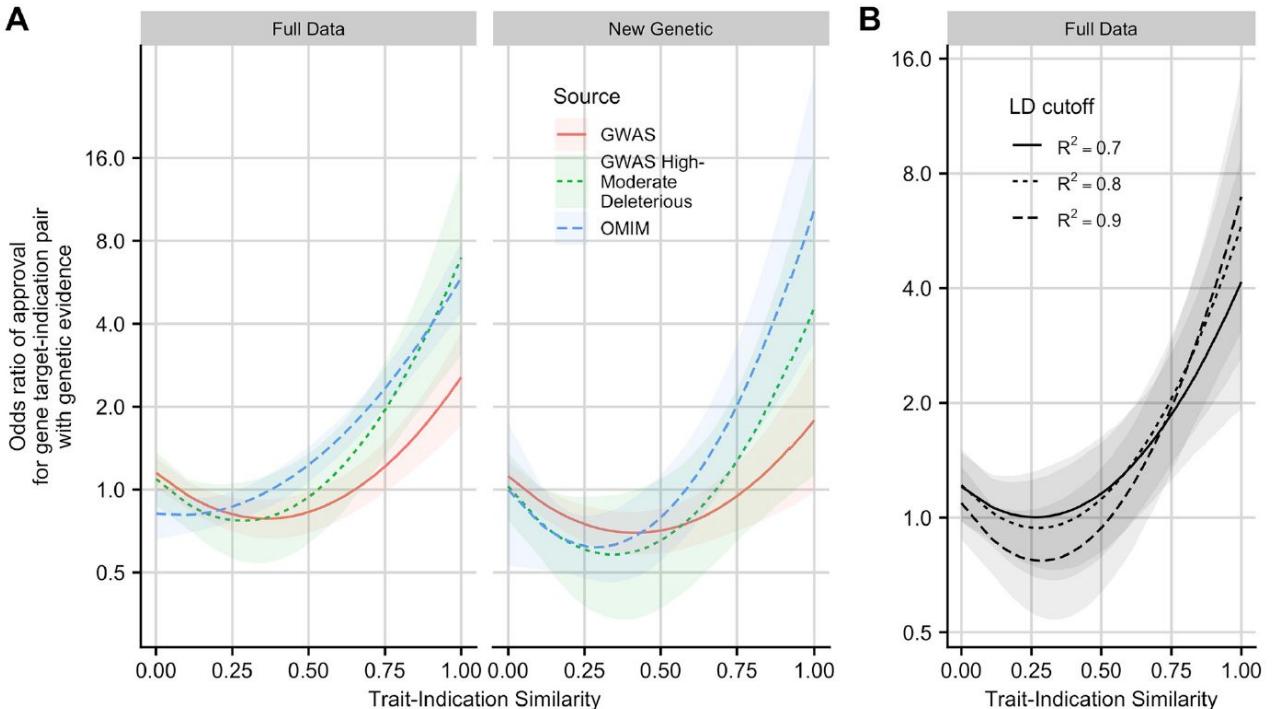


Fig 2. Estimated odds ratio of gene target-indication pair attaining approval, as a function of similarity between drug indication and the most similar trait associated with the target. A: Left: All genetic associations. Right: Only genetic associations reported after 2013 download. B: Effect of LD expansion threshold R^2 on the estimated approval odds ratio of a drug gene target-indication pair supported by a GWAS high-moderate deleterious variant. Posterior median and pointwise 95% credible interval from Bayesian logistic regression.

Discussion

What other evidences can we use to increase the likelihood that a gene is a good drug target?

GOT-IT recommendations for target-disease linkage

Assessment blocks



AB1: target–disease linkage (human targets)

1. Is the target perturbation a cause or consequence of the human disease process?
2. Is the therapeutic relevance (such as human connection) of models used sufficiently high for decision-making?
3. Is the target expression pattern known (that is, within the anticipated patient population)?
4. Is the target manipulation process clinically relevant?
5. Is the read-out used to detect target-dependent processes disease-relevant?
6. Is the stimulus used to activate or influence target-dependent processes disease-relevant?
7. Are the biological consequences of an observed effect size known?

Public resources for target assessment

AB1: target–disease linkage (human targets)

1. Is the target perturbation a cause or consequence of the human disease process?
2. Is the therapeutic relevance (such as human connection) of models used sufficiently high for decision-making?
3. Is the target expression pattern known (that is, within the anticipated patient population)?
4. Is the target manipulation process clinically relevant?
5. Is the read-out used to detect target-dependent processes disease-relevant?
6. Is the stimulus used to activate or influence target-dependent processes disease-relevant?
7. Are the biological consequences of an observed effect size known?

- [OpenTargets](#)
- [Online Mendelian Inheritance in Man](#) (OMIM)

- Scattered in diverse information sources such as [Wikipedia](#) and literature

- Health: [GTEX](#), [The Human Protein Atlas](#)
- Disease: [Gene Expression Atlas](#), scattered

Public resources for target safety assessment

AB2: target-related safety (human targets)

8. Is the target selective and not genetically linked to other diseases (or phenotypes or organ systems)?
9. Is there prior knowledge on safety of the target or reported evidence for the role of the target in a known pathway and/or physiological process that may be harmful if disrupted?
10. Are in vitro or pharmacologically relevant animal models available for safety testing?
11. Do models used for safety testing translate well to humans?
12. Are safety biomarkers available and can adverse effects be monitored and/or predicted by safety biomarkers?
13. Is there sufficient confidence that a necessary safety window has been or can be established?
14. Is the disease life-threatening (at what stage of the disease is the target of relevance)?
15. Is the tissue distribution of the target known (in humans or in animals)?

- [Comparative Toxicogenomics Database \(CTD\)](#)

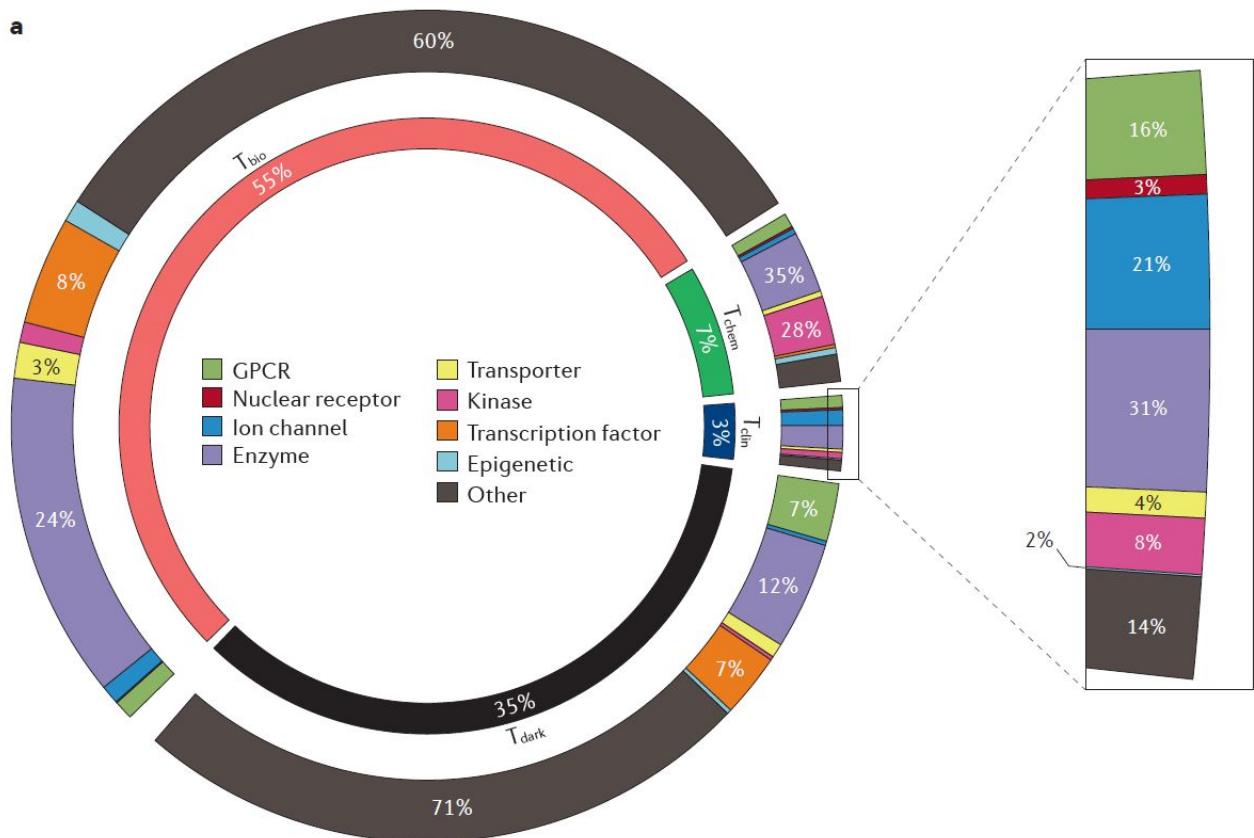
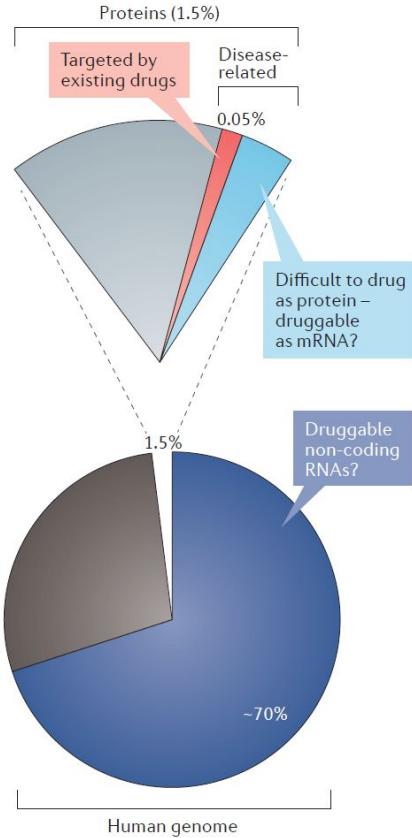
- [DrugBank, DrugCentral](#)
- [FDA Adverse Event Reporting System \(FAERS\)](#)

- [NCBI HomoloGene](#)
- [ENSEMBL ComparaGenom](#)
- [Mouse Genome Informatics \(MGI\)](#)

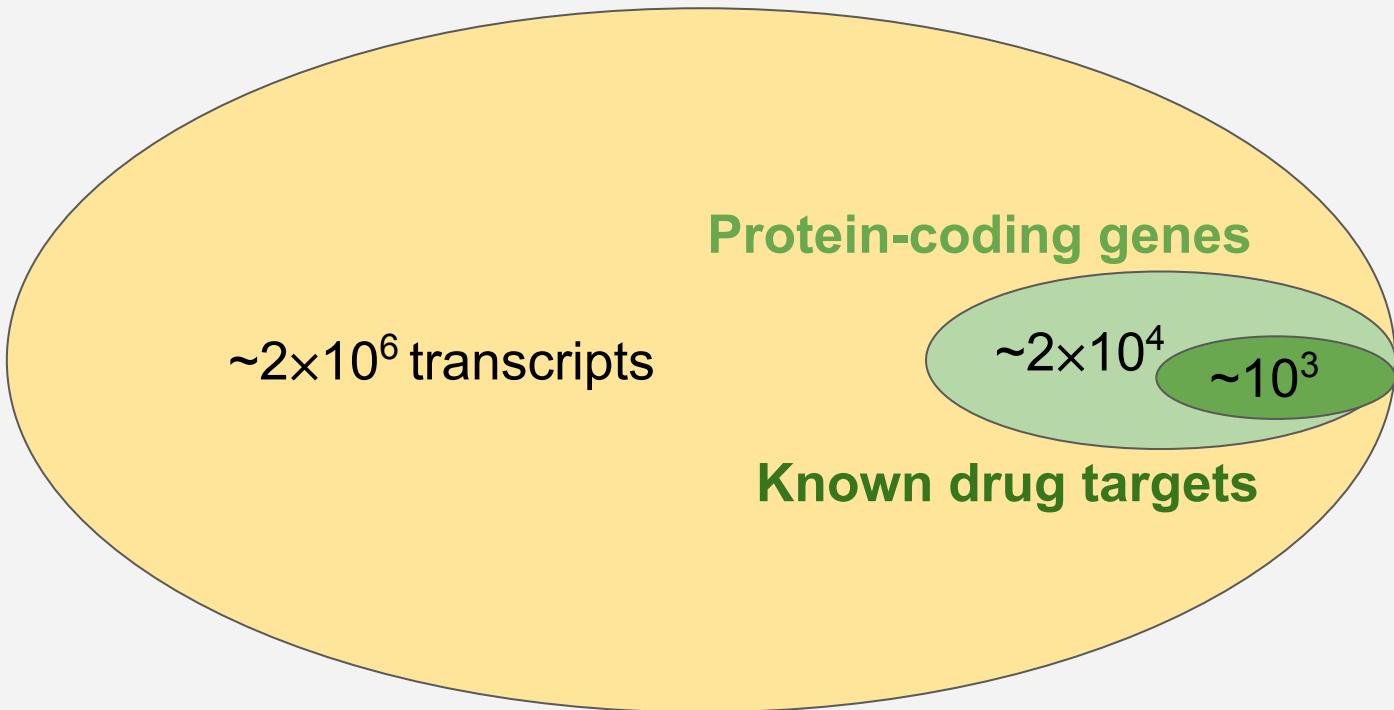
Other important information resources

- **Genomic variations:** [gnomAD](#), [dbSNP](#), and [TCGA](#) for oncology;
- **Protein domain and static structure:** [InterPro](#), [Pfam](#), and [PDB](#);
- **Interaction network and pathway:** [BioGRID](#), [IntAct](#), [Reactome](#), and [KEGG](#);
- **Gene expression profiles associated with the target:** [NCBI GEO](#) (Gene Expression Omnibus), [ARCHS4](#)

Challenge #1: little experience for much of the genome

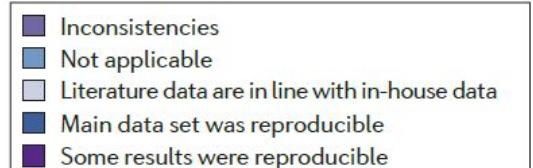
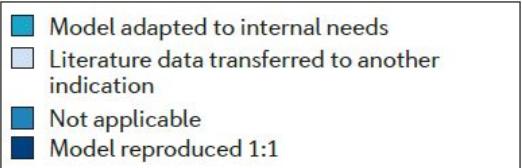
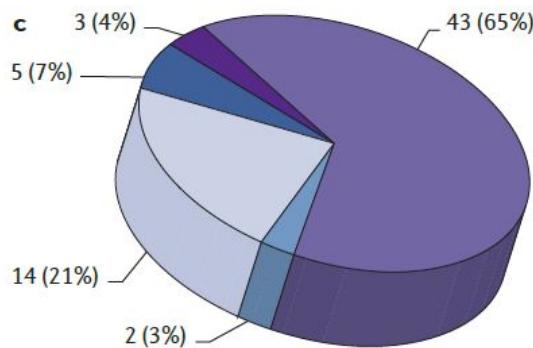
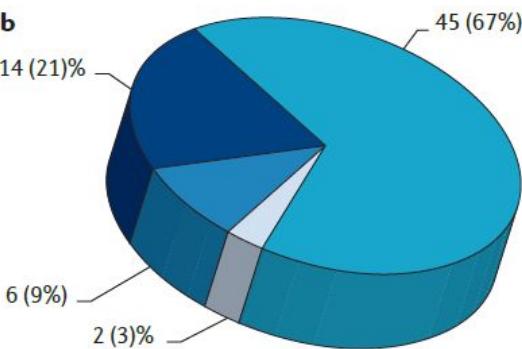
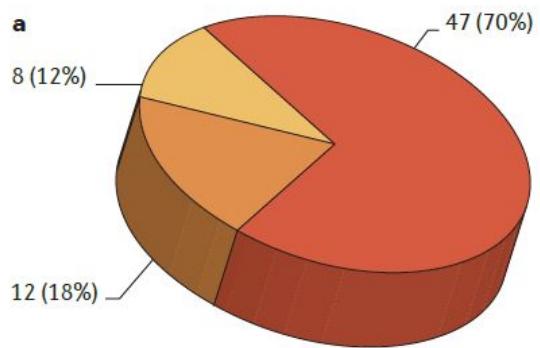


Protein, RNA, or DNA as target?



$\sim 3 \times 10^9$ DNA bases from maternal and paternal each

Challenge #2: Lack of reproducibility

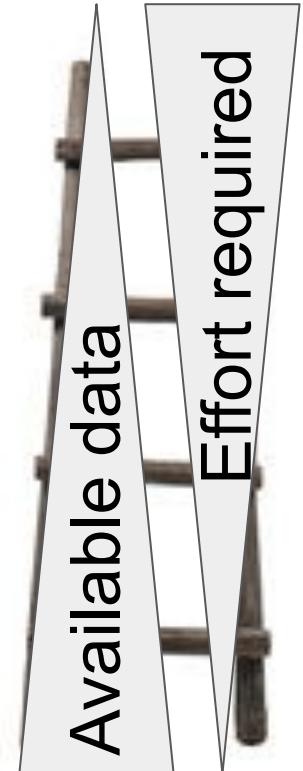


d

	Model reproduced 1:1	Model adapted to internal needs (cell line, assays)	Literature data transferred to another indication	Not applicable
In-house data in line with published results	1 (7%)	12 (86%)	0	1 (7%)
Inconsistencies that led to project termination	11 (26%)	26 (60%)	2 (5%)	4 (9%)

Challenge #3: The Target Ladder

3. [Real-world test] What would happen if we inhibit the activity of the kinase domain in the *WKN3* gene in patients with Alzheimer's Disease?
2. [Intervention] What would happen to human cells or to a rat model if we inhibit the activity of the kinase domain in the *WKN3* gene?
1. [Association] What are the evolutionary conservation, sequence, expression profile, expression regulation patterns, ... of the *WKN3* gene?



Conclusions

- Genomics and genetics offer unprecedented opportunities and challenges for target identification and assessment;
- Target identification and assessment involves knowledge integration and experimental validation;
- A central task of mathematical and computational biology in drug discovery is to perform inference, i.e. using information to reduce uncertainty.

References

1. Overington, John P., Bissan Al-Lazikani, and Andrew L. Hopkins. 2006. "How Many Drug Targets Are There?" *Nature Reviews Drug Discovery* 5 (12): 993–96. <https://doi.org/10.1038/nrd2199>.
2. Santos, Rita, Oleg Ursu, Anna Gaulton, A. Patrícia Bento, Ramesh S. Donadi, Cristian G. Bologa, Anneli Karlsson, et al. 2017. "A Comprehensive Map of Molecular Drug Targets." *Nature Reviews Drug Discovery* 16 (1): 19–34. <https://doi.org/10.1038/nrd.2016.230>.
3. Oprea, Tudor I., Cristian G. Bologa, Søren Brunak, Allen Campbell, Gregory N. Gan, Anna Gaulton, Shawn M. Gomez, et al. 2018. "Unexplored Therapeutic Opportunities in the Human Genome." *Nature Reviews Drug Discovery* 17 (February): 317–32. <https://doi.org/10.1038/nrd.2018.14>.
4. Finan, Chris, Anna Gaulton, Felix A. Kruger, R. Thomas Lumbers, Tina Shah, Jorgen Engmann, Luana Galver, et al. 2017. "The Druggable Genome and Support for Target Identification and Validation in Drug Development." *Science Translational Medicine* 9 (383): eaag1166. <https://doi.org/10.1126/scitranslmed.aag1166>.
5. Plenge, Robert M., Edward M. Scolnick, and David Altshuler. 2013. "Validating Therapeutic Targets through Human Genetics." *Nature Reviews Drug Discovery* 12 (8): 581–94. <https://doi.org/10.1038/nrd4051>.
6. Prinz, Florian, Thomas Schlange, and Khusru Asadullah. 2011. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?" *Nature Reviews Drug Discovery* 10 (9): 712–712. <https://doi.org/10.1038/nrd3439-c1>.
7. Matsui, Masayuki, and David R. Corey. 2017. "Non-Coding RNAs as Drug Targets." *Nature Reviews Drug Discovery* 16 (3): 167–79. <https://doi.org/10.1038/nrd.2016.117>.
8. Setten, Ryan L., John J. Rossi, and Si-ping Han. 2019. "The Current State and Future Directions of RNAi-Based Therapeutics." *Nature Reviews Drug Discovery* 18 (6): 421–46. <https://doi.org/10.1038/s41573-019-0017-4>.
9. Warner, Katherine Deigan, Christine E. Hajdin, and Kevin M. Weeks. 2018. "Principles for Targeting RNA with Drug-like Small Molecules." *Nature Reviews Drug Discovery* 17 (8): 547–58. <https://doi.org/10.1038/nrd.2018.93>.
10. Donovan, Katherine A., Fleur M. Ferguson, Jonathan W. Bushman, Nicholas A. Eleuteri, Debabrata Bhunia, SeongShick Ryu, Li Tan, et al. 2020. "Mapping the Degradable Kinome Provides a Resource for Expedited Degrader Development." *Cell* 183 (6): 1714–1731.e10. <https://doi.org/10.1016/j.cell.2020.10.038>.

References (continued)

11. Shih, Hsin-Pei, Xiaodan Zhang, and Alex M. Aronov. 2017. "Drug Discovery Effectiveness from the Standpoint of Therapeutic Mechanisms and Indications." *Nature Reviews Drug Discovery*, October. <https://doi.org/10.1038/nrd.2017.194>.
12. Nguyen, Phuong A., David A. Born, Aimee M. Deaton, Paul Nioi, and Lucas D. Ward. 2019. "Phenotypes Associated with Genes Encoding Drug Targets Are Predictive of Clinical Trial Side Effects." *Nature Communications* 10 (1): 1–11. <https://doi.org/10.1038/s41467-019-09407-3>.
13. Ding, Li, Matthew H. Bailey, Eduard Porta-Pardo, Vesteinn Thorsson, Antonio Colaprico, Denis Bertrand, David L. Gibbs, et al. 2018. "Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics." *Cell* 173 (2): 305–320.e10. <https://doi.org/10.1016/j.cell.2018.03.033>.
14. Hon, Chung-Chau, and Piero Carninci. 2020. "Expanded ENCODE Delivers Invaluable Genomic Encyclopedia." *Nature* 583 (7818): 685–86. <https://doi.org/10.1038/d41586-020-02139-1>.
15. Moore, Jill E., Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shoresh, Jessika Adrian, Trupti Kawli, et al. 2020. "Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes." *Nature* 583 (7818): 699–710. <https://doi.org/10.1038/s41586-020-2493-4>.
16. Kim-Hellmuth, Sarah, Fran ois Aguet, Meritxell Oliva, Manuel Mu oz-Aguirre, Silva Kasela, Valentin Wucher, Stephane E. Castel, et al. 2020. "Cell Type-Specific Genetic Regulation of Gene Expression across Human Tissues." *Science* 369 (6509). <https://doi.org/10.1126/science.aaz8528>.
17. Uhl n, Mathias, Linn Fagerberg, Bj rn M. Hallstr m, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu,  sa Sivertsson, et al. 2015. "Tissue-Based Map of the Human Proteome." *Science* 347 (6220): 1260419. <https://doi.org/10.1126/science.1260419>.
18. Wang, Dongxue, Basak Eraslan, Thomas Wieland, Bj rn Hallstr m, Thomas Hopf, Daniel Paul Zolg, Jana Zecha, et al. 2019. "A Deep Proteome and Transcriptome Abundance Atlas of 29 Healthy Human Tissues." *Molecular Systems Biology* 15 (2): e8503. <https://doi.org/10.15252/msb.20188503>.
19. Wang, Mingcong, Christina J. Herrmann, Milan Simonovic, Damian Szklarczyk, and Christian von Mering. 2015. "Version 4.0 of PaxDb: Protein Abundance Data, Integrated across Model Organisms, Tissues, and Cell-Lines." *PROTEOMICS* 15 (18): 3163–68. <https://doi.org/10.1002/pmic.201400441>.
20. Emmerich, Christoph H., Lorena Martinez Gamboa, Martine C. J. Hofmann, Marc Bonin-Andresen, Olga Arbach, Pascal Schendel, Bj rn Gerlach, et al. 2021. "Improving Target Assessment in Biomedical Research: The GOT-IT Recommendations." *Nature Reviews Drug Discovery* 20 (1): 64–81. <https://doi.org/10.1038/s41573-020-0087-3>.

References (continued)

21. Cao, Chen, and John Moult. 2014. "GWAS and Drug Targets." *BMC Genomics* 15 (4): S5. <https://doi.org/10.1186/1471-2164-15-S4-S5>.
22. Patron, Jonas, Arnau Serra-Cayuela, Beomsoo Han, Carin Li, and David Scott Wishart. 2019. "Assessing the Performance of Genome-Wide Association Studies for Predicting Disease Risk." *PLOS ONE* 14 (12): e0220215. <https://doi.org/10.1371/journal.pone.0220215>.
23. <https://www.genome.gov/About-Genomics/Introduction-to-Genomics>
24. Jansen, An, and Kevin J. Verstrepen. 2011. "Nucleosome Positioning in *Saccharomyces Cerevisiae*." *Microbiology and Molecular Biology Reviews* 75 (2): 301–20. <https://doi.org/10.1128/MMBR.00046-10>.
25. https://gnomad.broadinstitute.org/gene/ENSG00000141510?dataset=gnomad_r2_1
26. Collins, Ryan L., Harrison Brand, Konrad J. Karczewski, Xuefang Zhao, Jessica Alföldi, Laurent C. Francioli, Amit V. Khera, et al. 2020. "A Structural Variation Reference for Medical and Population Genetics." *Nature* 581 (7809): 444–51. <https://doi.org/10.1038/s41586-020-2287-8>.
27. NCBI Gene of CFTR: <https://www.ncbi.nlm.nih.gov/gene/1080>
28. Hughes, Maria F., Yvonne M. Lenighan, Catherine Godson, and Helen M. Roche. 2018. "Exploring Coronary Artery Disease GWAs Targets With Functional Links to Immunometabolism." *Frontiers in Cardiovascular Medicine* 5. <https://doi.org/10.3389/fcvm.2018.00148>.
29. Consortium, The GTEx. 2020. "The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues." *Science* 369 (6509): 1318–30. <https://doi.org/10.1126/science.aaz1776>.
30. Baralle, Marco, and Francisco Ernesto Baralle. 2018. "The Splicing Code." *Biosystems, Code Biology*, 164 (February): 39–48. <https://doi.org/10.1016/j.biosystems.2017.11.002>.

References (continued)

32. https://en.wikipedia.org/wiki/Haemophilia_in_European_royalty
33. Smemo, Scott, Juan J. Tena, Kyoung-Han Kim, Eric R. Gamazon, Noboru J. Sakabe, Carlos Gómez-Marín, Ivy Aneas, et al. 2014. "Obesity-Associated Variants within FTO Form Long-Range Functional Connections with IRX3." *Nature* 507 (7492): 371–75. <https://doi.org/10.1038/nature13138>.
34. Nelson, Matthew R., Hannah Tipney, Jeffery L. Painter, Judong Shen, Paola Nicoletti, Yufeng Shen, Aris Floratos, et al. 2015. "The Support of Human Genetic Evidence for Approved Drug Indications." *Nature Genetics* 47 (8): 856–60. <https://doi.org/10.1038/ng.3314>.
35. Ardlie, Kristin G., Leonid Kruglyak, and Mark Seielstad. 2002. "Patterns of Linkage Disequilibrium in the Human Genome." *Nature Reviews Genetics* 3 (4): 299–309. <https://doi.org/10.1038/nrg777>.
36. Brennan, Richard J. 2017. "Target Safety Assessment: Strategies and Resources." In *Drug Safety Evaluation: Methods and Protocols*, edited by Jean-Charles Gautier, 213–28. *Methods in Molecular Biology*. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-7172-5_12.

An example of complementary views

We want to work on hepatocarcinoma (liver cancer) and have the following information about a potential target X:

- X is a receptor expressing on the surface of most cell types;
- Upon binding ligands, X activates innate immune response;
- Gene sequence of X is conserved in primates but *not* in rodents;
- Protein X interacts with protein Y, which is essential, namely Y knockout causes lethal embryos;
- Asian population has a unique genetic variant in the non-coding region of X;

Discussion: what are the consequences of having these information?